Chapter 1

1. Introduction

The origin of the theory of discrete probability distributions began with the works of James Bernoulli and Poisson. The Swiss mathematician, James Bernoulli derived the Binomial distribution and published it in the year 1713. Poisson distribution was derived by a French mathematician Simeon D. Poisson as limiting form of the binomial distribution in 1837. In 1920, Greenwood and Yulo obtained negative Binomial distribution as a consequence of certain assumptions in accident proneness models.

It is difficult or inconvenient to get samples from continuous distributions. In real life situation the observed values are discrete in nature; they are measured only in a finite number. Even if the measurements are taken in continuous scale the observations may be recorded in a way making discrete model more appropriate. It is therefore reasonable to consider the observations as coming from a discrete distribution generated from underlying continuous model. Discrete distributions obtained by discretizing a continuous failure time model have appeared in the statistical literatures.

It is sometimes impossible or inconvenient to measure the life length of a device, on a continuous scale. In practice, we come across situations, where lifetime of a device is considered to be a discrete random variable. For example, in the case of an on/off- switching device, the lifetime of the switch is a discrete random variable. If the lifetimes of individuals in some population are grouped or when lifetime refers to an integral number of cycles of some sort, it may be desirable to treat it as a discrete random variable. When a discrete model is used with lifetime data, it is usually a multinomial distribution. This arises because effectively the continuous data have been grouped. Some situations may demand another discrete distribution, usually over the non-negative integers. Such situations are best treated individually, but generally one tries to adopt one of the standard discrete distributions.

1.1 Background of study

1.1.1 Discretization of Continuous Distributions

Discretization is a popular approach to handling numeric attributes in machine learning. Discretization methods can be classified as equal width discretization (EWD), equal frequency discretization (EFD), fuzzy discretization (FD), entropy minimization discretization (EMD), iterative discretization (ID), proportional k-interval discretization (PKID), lazy discretization (LD), non-disjoint discretization (NDD) and weighted proportional k-interval discretization (WPKID). Discretization of continuous random variables is very useful. An example is when you need to compute the distribution of a compound random variable. This is the case in actuarial science, where the aggregate claims are sum of individual claims, and the number of individual claims is itself a random variable.

1.1.2 Size-biased distributions

Size-biased distributions arise naturally in practice when observations from a sample are recorded with unequal probabilities, having probability proportional to size (PPS). It is a more general form known as weighted distributions. The most common cases of size- biased distribution occur when $\alpha = 1, or\ \alpha = 2$, these special cases may be term as length and area biased respectively. If a discrete random variable X having pmf $f(x;\theta)$, hence the pmf of the size-biased distribution may be obtained as

$$f^*(x;\theta) = \frac{x\,\mathrm{f}(x;\theta)}{\mu'},$$

where $E(X) = \mu'$, is the mean of the distribution.

1.1.3 Zero Modified Discrete Probability Distributions

A counting distribution is a discrete distribution with only non-negative integers in its domain. We typically use a counting distribution to model the number of occurrences of a certain event, for example "number of car accidents in a year". The Poisson distribution is a counting distribution for example. However, it can occur that "zero events occurring" is not properly modeled by the counting distribution. A solution for this is to use a zero-modified distribution, which alters the probability of occurrence of zero. There are basically two ways of modifying the counting distribution at zero:

1.1.4 Zero Truncated Discrete Probability Distributions

A truncated distribution is a conditional distribution that results from restricting the domain of some other probability distribution. Truncated distributions arise in practical statistics in cases where the ability to record, or even to know about, occurrences is limited to values which lie above or below a given threshold or within a specified range. For example, if the dates of birth of children in a school are examined, these would typically be subject to truncation relative to those of all children in the area given that the school accepts only children in a given age range on a specific date. There would be no information about how many children in the locality had dates of birth before or after the school's cutoff dates if only a direct approach to the school were used to obtain information.

1.1.5 Estimation of Parameter

Estimation theory is a branch of statistics that deals with estimating the values of parameters based on measured data that has a random component. The parameters describe an underlying physical setting in such a way that their value affects the distribution of the measured data. An estimator attempts to approximate the unknown parameters using the measurements. For example, it is desired to estimate the proportion of a population of voters who will vote for a particular candidate. That proportion is the parameter sought; the estimate is based on a small random sample of voters.

1.1.6 Review of Literature

There are many papers related to the discrete distributions have been published in various journals. A brief review of literatures related to my topic of investigation are given below:

The paper by Borah and Begum [7] discussed on some properties of Poisson–Lindley distribution and its derived distributions. The paper by Borah and Deka Nath [8] discussed on Poisson-Lindley distribution and some of its Mixture distributions. Also they studied on certain Inflated Poisson Lindley distributions.

Discrete geometric distribution can be obtained by discretizing the exponential continuous distribution. Some of those works of Nakagawa and Osaki [35]; where the discrete Weibull distribution is obtained. Roy [39] studied the discrete Rayleigh distribution.

Gómez-Déniz and Calderin-Ojeda [22] derived a new generalization of the geometric distribution obtained from the generalized exponential distribution of Marshall and Olkin [33]. They has also derived discrete Lindley distribution.

Khan et al. [27] estimated the parameter of discrete Weibull distributions. Krishna and Pundir [28] derived the discrete Burr distribution.

Sato et al. [41] proposed discrete exponential and Gamma distribution and its application. Chakraborty and Charkravorty [10] also derived discrete gamma distribution for modeling real life data.

Zakerzadeh and Dolati [54] generalized Lindley distribution and introduced a three–parameter generalized Lindley distribution which includes the exponential and gamma distributions as a special case.

Ghitany et. al. [17, 18, 19, 20] have discussed various properties of Lindley distribution.

1.2 Methodologies/approach(es) applied

1.2.1 Discretization of continuous distribution

If the underlying continuous failure time X has the survival function $S(x) = \Pr(X > x)$, the probability mass function of the discrete random variable associated with that continuous distribution can be obtained as

$$\Pr(X = x) = S(x) - S(x + 1), \ x = 0, 1, 2, \ldots$$

1.2.2 Size-biased discrete distribution

Size-biased distributions arise naturally in practice when observations from a sample are recorded with unequal probabilities, having probability proportional to size (PPS). It is a more general form known as weighted distributions. Fisher [16] first introduced these distributions to model ascertainment bias which were later formalized by Rao [38]. If the random variable X has distribution $f(x; \theta)$, with unknown parameter θ, then the corresponding weighted distribution is of the form

$$f^{w}(x; \theta) = \frac{w(x) f(x; \theta)}{E[w(x)]},$$

where $w(x)$ is a non-negative weight function such that $E[w(x)]$.

A special case of interest arises when the weight function $w(x) = x^{\alpha}$. Such distributions are known as size-biased distributions of order α. The most common cases of size-biased distribution occur when $\alpha = 1, or\ \alpha = 2$, these special cases may be term as length and area biased respectively. If a discrete random variable X having pmf $f(x; \theta)$, hence the pmf of the size-biased distribution may be obtained as

$$f^{*}(x; \theta) = \frac{x f(x; \theta)}{\mu'},$$

where $\mu' = E(X)$, is the mean of the distribution.

Van Deusen [51] discussed size-biased distribution theory and applied it to fitting distributions of diameter at breast height (DBH) data arising from horizontal point sampling (HPS). Later, Lappi and Bailey [29] used size-biased distributions to analyze HPS diameter increment data. Most of the statistical applications of these distributions, especially to the analysis of data relating to human population and ecology can be found in Patil and Rao [36]. Gove [23] reviewed some of the recent results on size-biased distributions pertaining to parameter estimation in forestry, with special emphasis on Weibull family.

1.2.3 Zero Truncated Discrete Probability Distributions

If we remove the probability of zero (P(0)) occurring then we speak of a zero-truncated model

$f(x) = \pi * P(x)$,
where, $\pi = 1/(1 - P(0))$.

There is a variety of applications, e.g. in accident or family studies where certain values are missing or may not be recorded; hence, truncated distributions are used to explain the model underlying the truncation process. A general method of study may be used based on the derivation of the probability generating function (pgf) of the truncated distributions. It is noted that in all cases, the pgf's of the truncated distributions are simple linear function of the pgf of the initial distribution. That linear relation enables us to derive the various properties of the truncated version of the distribution. There are three types of truncation. For both left truncation and right truncation problems, variety of special cases can be presented.

1.2.4 Zero Modified Discrete Probability Distributions

A probability distribution is called zero-inflated when it contains an excessive number of zeroes. The estimations of any basic parametric distributions under estimate the probability mass at zero. Alternatively, zero-inflated distributions can be used. These distributions belong to a finite mixture family. That is, a zero-inflated distribution consists of two processes. One process process generates Zeroes while the other process generates zeroes and nonzeroes.

Zero-inflated (discrete) distributions are distributions with an extra probability assigned to zero as an outcome. Suppose we have a distribution with probability mass function (x) , and let $0 < w \leq 1$, then we define the probability mass of the Zero-Inflated (ZI) distribution as

$$P(0) = (1 - w) + w.f(0)$$
$$P(x) = w.f(x) \quad \text{for} \quad x = 1, 2, 3, \ldots$$

1.2.5 Probability generating Function

Consider a nonnegative discrete rv X with nonzero probabilities only at nonnegative integer values. Let

$$P_x = \Pr[X = x], \qquad x = 0,1,2 \ldots$$

If the distribution is proper, then $\sum_{x=0} P_x = 1$ and hence $\sum_{x=0} P_x\, t^x$ converges for $|t| \leq 1$ The probability generating function of the distribution with probability mass function is defined as

$$G(t) = \sum_{x=0}^{\infty} P_x t^x = E[t^x].$$

1.2.6 Factorial moment generating function (fmgf)

The corresponding fmgf for the rv X taking the values 0,1,...,k… with probabilities $P_0, P_1, P_2 P_3, \ldots$, is defined as

$$M(t) = G(1 + t) = \sum_{x=0}^{\infty} P_x (1 + t)^x.$$

The rth factorial moment is obtained from moment as

$$\mu'_{[r]} = \left[\frac{d^r M(t)}{dt^r}\right]_{t=0} .$$

1.2.7 Survival function

The survival function for continuous probability density function is defined as

$$S(x) = \int_{t=x}^{\infty} P_t\, dt.$$

1.2.8 Cumulative function

The cumulative function for continuous probability density is defined as

$$F(x) = \int_{t=o}^{x} P_t \, dt.$$

1.2.9 Survival function

The survival function for discrete probability density function is defined as

$$S(x) = \sum_{t=x}^{\infty} P_t.$$

1.2.10 Cumulative function

The cumulative function for discrete probability density is defined as

$$F(x) = \sum_{t=0}^{x} P_t \, .$$

1.2.11 Failure hazared rate function

The failure hazared rate function is defined as

$$r(x) = \frac{P(X=x)}{P(X \geq x-1)} \, .$$

1.2.12 Reversed failure rate function

The Reversed failure function is defined as

$$r^*(x) = \frac{P(X=x)}{P(X \leq x)} \, .$$

1.2.13 The second rate of failure function

The second rate of failure is defined as

$$r^*(x) = log\left[\frac{s(x)}{s(x+1)}\right].$$

1.3 Estimation

1.3.1 Method of Maximum Likelihood

The method of maximum likelihood is widely advocated. Let x_1 , x_2 , x_3,…, x_n be a random sample of size n from a population with probability function $P_x(\theta_1, \theta_2, \dots, \theta_m)$ and let f_x bethe observed frequency in the sample corresponding to $X = x,\ (x = 1, 2, \dots, k)$ such that $\sum_{x=1}^{k} f_x = n$, where k is the largest observed value having non–zero frequency. The likelihood function, L of the probability function $P_x(\theta_1, \theta_2, \dots, \theta_m)$ is given by

$$L = \prod_{x=1}^{k} {P_x}^{f_x} .$$

Maximizing the Likelihood can usually be achieved by solving the equations

$$\frac{\partial L}{\partial \theta_1} = 0.$$

$$\frac{\partial L}{\partial \theta_2} = 0.$$

$$\frac{\partial L}{\partial \theta_m} = 0.$$

These equations are called the maximum-likelihood equations.

1.3.2 Method of Moments

It consists in equating the moments of the population to the sample moments and then to solving the equations so obtained to get the required estimates of the population parameter.

1.4 Objectives

The objectives of the present study are

- Generalization of certain discrete distributions and study of its properties.
- Discretization of certain continuous distributions.
- Derivation of Size-biased discrete distributions.
- Derivation of Zero-Modified discrete probability distributions.

- Derivation of Zero-Truncated discrete probability distributions.
- Estimation of parameters and fitting of the distributions.

1.5 Out line of the thesis

Chapter 2. A two parameter continuous distribution named "Janardan distribution (JD)", introduced by Rama Shanker, Shambhu Sharma, Umma Shanker and Ravi Shanker [48] has been considered. They derived certain properties of continuous Janardan distribution and also discussed stochastic orderings. They also discussed maximum likelihood method and method moment for estimating its parameter.

Our objective is to discretize the continuous Janardan distribution. A new discrete probability distribution called discrete Janardan (DJ) distribution has been derived. Its pgf and fmgf failure rate function have been derived. Zero truncated form of this distribution has been derived and certain properties of the distribution have also been studied. The parameters of the distribution have been estimated using the method of maximum likelihood and method of moments for the fitting of the distribution

Chapter 3. A two parameter continuous Sushila Distribution has been introduced by Rama Shanker, Shambhu Sharma, Umma Shanker and Ravi Shanker [47]. They derived its moments failure rate functions mean residual function and discussed stochastic orderings. They also discussed maximum likelihood method and method moment for estimating its parameter.

Continuous Sushila distribution has been discretized. The new discrete probability distribution may be called discrete Sushila (DS) distribution. The pgf and fmgf, failure rate function have been derived. Zero truncated form of this distribution has also been obtained. The method of maximum likelihood and method of moment have been used for the estimation of its parameters.

Chapter 4. A two-parameter Quasi Lindley distribution (QLD), introduced Rama Shanker and Mishra [44]. They derived its moments, failure rate function, mean residual life function and stochastic orderings. The maximum likelihood method and the method of moments have been discussed for estimating its parameters. The distribution has been fitted to some data-sets for testing its goodness of fit.

Two parameter Quasi Lindley distribution has been discretized which may be called it discrete Quasi-Lindley distribution. The pgf and fmgf have been derived. The recurrence relations have been also derived respectively. Failure rate functions, zero truncated, size biased, zero modified form the discrete Quasi Lindley distribution are obtained. Maximum Likelihood methods, method of moment are used to find in parameter estimation.

Chapter 5. A new two parameters Quasi Lindley distribution (NQLD), introduced by Rama Shanker and Amanuel Habte Ghebretsadik [42]. They derived its moments, failure rate function, mean residual life function and stochastic orderings. The maximum likelihood method and the methods of moments have been discussed for estimation of parameters. The distribution has been fitted to some data-sets to test its goodness of fit.

Here we discretized the continuous NQDL distribution and named it Two parameter discrete Quasi Lindley distribution. The probability generating function, factorial moment generating function, recurrence relations are derived respectively. Failure rate functions, zero truncated, size biased, zero modified form the Discrete Quasi Lindley distribution have been obtained. Maximum Likelihood methods, method of moment are used to find in parameter estimation.

Chapter 6. Two-parameter Lindley distribution, introduced by Rama Shanker, et. al. [46]. They derived its statistical properties including its shape, moments, skewness, kurtosis, hazard rate function, mean residual life function, stochastic ordering, mean deviations, order statistics, Renyi entropy measure, Bonferroni and Lorenz curves, stress-strength reliability. For estimating its parameters, maximum likelihood estimation has been discussed. Finally, a numerical example has been presented to test the goodness of fit of the proposed distribution and the fit has been compared with the two-parameter generalized Lindley distribution.

We discretized the continuous two parameter Lindley distribution. The derived distribution may be called Two Parameter Discrete Lindley distribution (TPDL) and derived probability generating function, factorial moment generating function. The recurrence relations are also derived respectively. Failure rate functions, zero truncated, size biased, zero modified form the TPDL distribution are obtained. Maximum Likelihood method, method of moment have been used to find in parameter estimating its parameter.

Chapter 7. Two parameter continuous distribution gamma-Lindley distribution has been introduced by H. Zeghdoudi and S. Nedjar [53] in the paper "On gamma Lindley distribution". They derived its moments failure rate functions mean residual function and discussed stochastic orderings. They also discussed maximum likelihood method and method moment for estimating its parameter.

Here we discretized the continuous gamma-Lindley distribution and named it discrete gamma Lindley distribution. The probability generating function, factorial moment generating function, the recurrence relations are also derived respectively. Failure rate functions, zero truncated, size biased, zero modified form the discrete gamma Lindley distribution are obtained. Maximum Likelihood methods, method of moment have been used for the estimation of parameters.